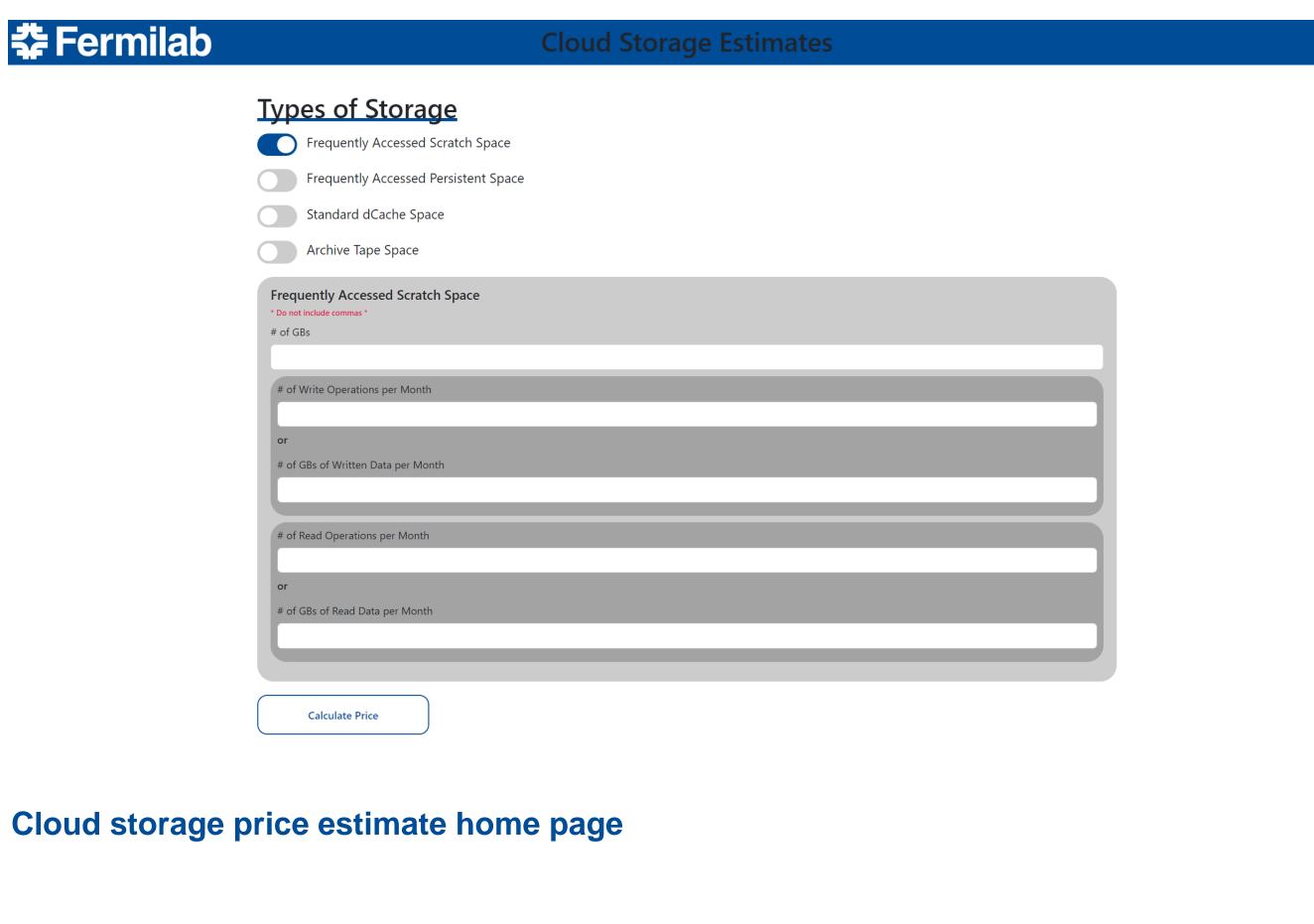
Understanding and Presenting Scientific Data Storage Usage

Alex Knusel, OMNI Intern, University of Minnesota

Introduction

The idea behind this project was to find a way for groups at Fermilab using Fermilab's scientific data storage services to better understand the amount of data and resources they are using. Multiple ways were brainstormed to determine the optimal way to achieve this, but the approach taken was to give the storage a monetary value. Fermilab does not charge individual groups usage fees for their storage except for Archive Facility users. Instead, adding hypothetical costs to the service would allow for more mindful storage usage. Research was conducted on multiple cloud computing services such as Google Cloud, Amazon Web Services, IBM, and Microsoft Azure to determine how to put a price tag on Fermilab's cloud storage. This information was collected, and a website using Flask was created to calculate, display, and inform users about their group's cloud storage usage.



	Microsoft Azure				Google Cloud					IBM Cloud Object Storage				Amazon Cloud Storage (S3)	
Plans	Hot Tier	Cool Tier	Archive	Plans	Standard	Nearline	Coldline	Archive	Plans	Smart	Standard	Vault	Cold Vault	Notes Pricing	S
Frequency	Whenever	30 day min	180 day min	Frequency	Whenever	30 day min	90 day min	365 day mir	Frequency	Whenever	Frequently	Monthly	A Few Times Yearly	Amazon has a	lot of good organized
Storage Size	rage Size			Storage Size	Unlimited Storage Size				Storage Size	Unlimited			information on their site. Also, they have		
Storage Cost (per GB)				Storage Cost (per GB)	\$0.020	\$0.010	\$0.00	4 \$0.00	12 Storage Cost (per GB)					their own built	in price calculator so I
First 50 TB / month	\$0.021	\$0.015	\$0.00099						Under 500 TB		\$0.022	\$0.012	\$0.006	thought it would	be redundant making or
Next 450 TB / month	\$0.02	\$0.015	\$0.00099						Over 500 TB		\$0.02	\$0.01	\$0.006	on here. I have h	ad some experience wit
Over 500 TB / month	\$0.0191	\$0.015	\$0.00099						Smart Only					using Amazon in the past with hosting VN	
Yearly Plans						Hot		\$0.021				and it was very reliable.			
1 Year reserved (100 TB /															
month)	\$1,545	\$840	\$91						Cool	\$0.012					
3 years reserved (1 PB /															
month)	\$15,050	\$8,179	\$883						Cold	\$0.0075					
•	. ,	,							Operation Charges	·					
Operation Charges (per				Operation Charges (per					(per 10,000						
10,000 operations)				10,000 operations)					operations)						
Write operations	\$0.07	\$0.13	\$0.13	Class A operations	\$0.05	\$0.10	\$0.1	0 \$0.	50 Class A operations	\$0.005	\$0.005	\$0.010	\$0.025		
Read operations	\$0.01			Class B operations	\$0.004		-	-	50 Class B operations	\$0.005					
Iterative read operations	\$0.01				,	,		•	Data retrieval (per GB)		None	\$0.01			
	,	,	******						Public Outbound				,		
Iterative write operations	\$0.07	\$0.13	\$0.13						Bandwidth						
Data retrieval	N/A		· · · · · · · · · · · · · · · · · · ·						0-50 TB	\$0.09	\$0.09	\$0.09	\$0.09		
Delete		Free	, , , , , , , , , , , , , , , , , , ,						Next 100 TB	\$0.07	-		-		
Data write		Free							Next 350 TB	\$0.05	-				
Index (GB/month)	\$0.03 N/A N/A							500+ TB	,		ontact IBM				
All other operations	\$0.01														
# of GBs	550,000			# of GBs	100,00	100,000	0 100,0	00 100.0	000 # of GBs		50,000	50,000	50,000		
Cost (Calculated based on				Cost (Calculated based					Cost (Calculated based	1					
# of GBs)	\$ 11,005.00	\$ 1,500.00	\$ 99.00	on # of GBs)	\$ 2,000.00	\$ 1,000.00	\$ 400.0	0 S 120.	00 on # of GBs)	•	\$ 1,100.00	\$ 600.00	\$ 300.00		
Notes	Ų 11/000100	ψ 1,500.00	φ 33.00	0.1.11 0.1 0.007	Ç 2,000.00	ψ 2,000.00	, , , , , , , , , , , , , , , , , , , ,	y 120.	Smart Hot GB's	50,000	,	Ç 000.00	ψ σσσισσ		
10123									Sindic flot GB 5	30,000	<u> </u>				
	Driging also can shange based on the type of											Sum of Star	ndard, Vault, and Cold		
Pricing also can change based on the type of redundency, but the cost doesn't change								Smart Cool GB's	50,000)	Vault Prices (Vault Prices (for Smart comparison)			
								Smart Cold GB's	50,000)	\$ 2,000.00				
		drastically							Smart Cost	\$2,025.0000					
Average Archive:	\$0.00392														
Average Standard:	\$0.015														

Information from different cloud computing services

Research

The bulk of the research for this project was conducted on other cloud computing and storage services such as Google Cloud, Amazon Web Services, IBM, and Microsoft Azure. The primary information collected from these services was the different plans and pricing they offered to users for using their services. This data was collected and put into a spreadsheet to look for similarities among the companies' services. The main similarities came from three different tiers that the services offered. These tiers were differentiated based on the frequency of data storage and access, and the amount of storage needed. The average prices were taken from each tier and then applied to the respective type of storage Fermilab offers for scientific data storage. Using data already stored, formulas for different types of storage were used to calculate previous months and years usage costs.

Website

The website was created using a Python Library called Flask which is a web framework with a small easy-toextend core that allows you to develop web applications. The main reason for using Flask instead of using just HTML, CSS, and JavaScript was the ability to update the webpage live with users' storage data. The Fermilab scientific data storage servers store data about the amount of storage each storage group uses for each type of storage. Using Flask, the website displays the information from the data file already generated, allowing for the storage groups to easily see their usage while making the website easy to update. This is the most important feature of the website because it allows the storage groups to see the amount of storage they are using and the hypothetical cost of the storage. This will help storage groups see the impact and cost of the information they are storing allowing for more mindful storage usage. While also helping the service provider to provide storage services more efficiently.



Individual group storage usage webpage

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

